Metrics & Terminology for Identification System Performance addressing Watch-list and other open set applications

Tony Mansfield

Tony.Mansfield@NPL.CO.UK

International Biometric Performance Conference   2 March 2010

Outline

1. Introduction & background

– Aspects of identification performance not covered by standards *ISO/IEC19795 Biometric performance testing and reporting*

2. Terminology

– Are existing terms sufficient for the variety of identification applications?

3. Metrics

– Image-level

– Subject-level

– Operator-level

4. Implications for test methodology

Background: Practical experience

**Face recognition identification against watch-list (2004)**

- CCTV set up in corridor

- Good quality face image photos for watch-list subjects

- Non-cooperative subjects - normal daily business - no operator involvement

- Ground truth established by RFID tag (with staff pass)

- Data collection over several days

  – RFID recognitions of test subjects

  – Face recognition candidate lists and probe image (if over threshold)

  – Presence and direction of travel count using IR beam

- Exploratory tests – what affects performance levels?

NPL
National Physical Laboratory

Background: Test standards

19795-1: Principles and framework

- Focus is on verification performance
    - Well established ground-truth
    - Cooperative test subjects
- Covers identification when the transaction is similar to that typical for verification
- Does not fully address complex identification applications
    - Example CCTV watch-list applications
        - Operator involvement
        - Multiple samples per subject
        - Overlapping between subjects
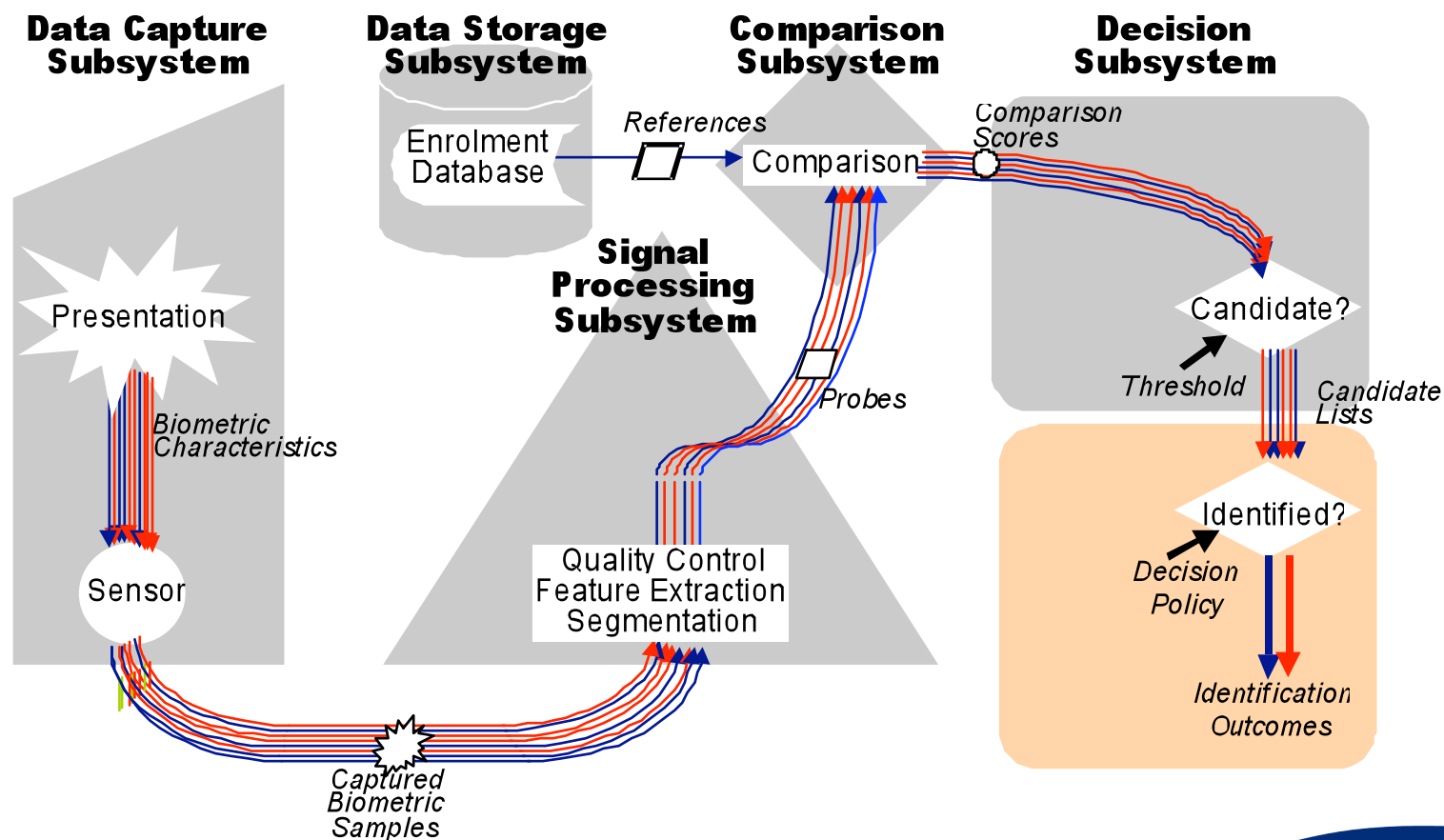   - Multiple subjects per image

19795-1: General biometric system
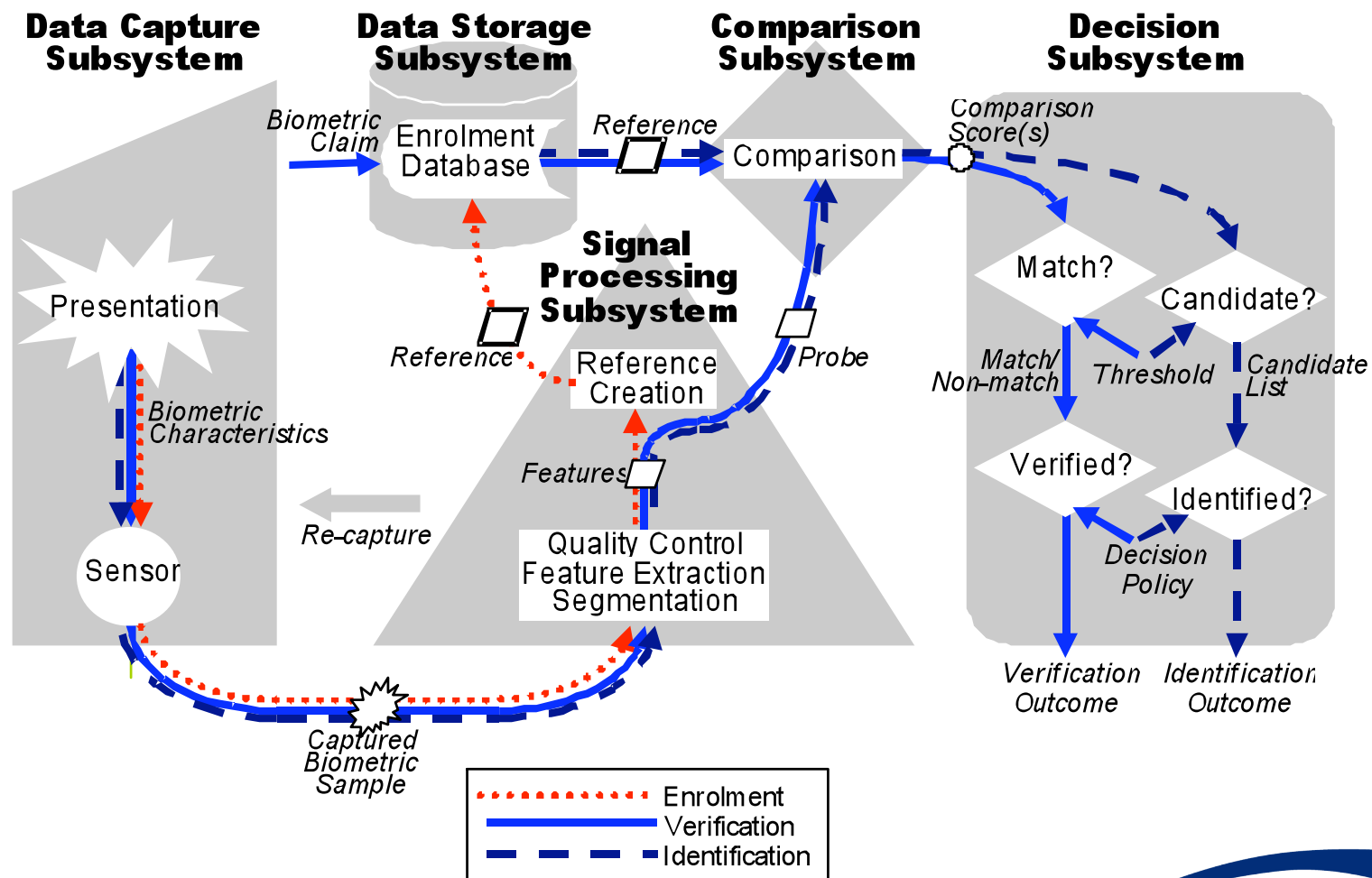
19795-1: Model of biometric identification system
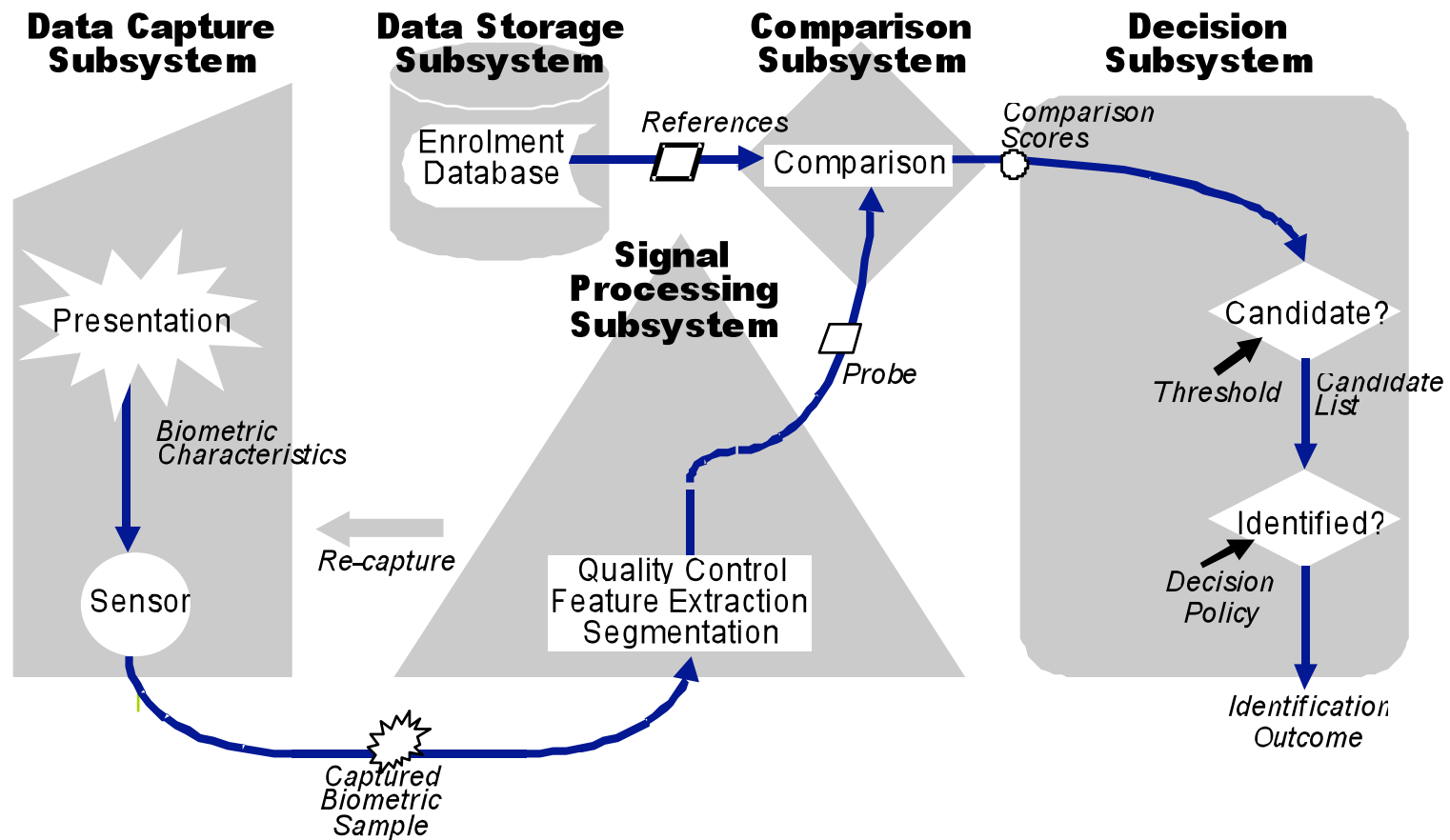
19795-1: Complex model of biometric identification system

Metrics for identification performance

Metrics from 19795-1 … Framework and Principles

- Failure to acquire

    – Do we have an image of the subject, and is it usable?

- Pre-selection error rate & penetration rate

    – Applies to binning / filtering algorithms,

    – Can be applied at final stage of refining the candidate list too (Reliability & Selectivity)

- False-positive & false-negative identification-error rates

    – Identification decision errors

    – Make assumptions on how the candidate list will be used to inform the identification decision

NPL
National Physical Laboratory

Identification errors?

If system requires candidate list of the top 10 matches

- Not an error to return matches for a subject not in the watch-list?

Correctness depends on the comparison scores assigned

- Not an error to ascribe incorrect identity if correctly assigned low probability

- E.g. DNA: proportion of population having an equal or better match between probe and reference DNA sample

- Not an identification error to fail to ascribe an identity when insufficient biometric evidence

Correctness may depend on application

- Not necessarily an to ascribe incorrect identity to subject known to the system if the resulting outcome is correct

Identification metrics:
Image viewpoint

- Evaluated over BioAPI_IdentifyMatch calls

- Using the (segmented) image as a probe, is the correct reference returned in the candidate list.

- c.f. FMR and FNMR for verification

- Pros:
  - Easy to evaluate

- Cons
  - Does not show results of best match
  - Subjects where no segmented image acquired not represented in results
  - Biased by image segmentation (face finder) algorithms used

Identification metrics:
Subject viewpoint

- Evaluated over instances where test subject is present at the identification station

- Using full sequence of images obtained, what is probability of detection if on watch-list, or probability of false alarm if not on watch-list

- c.f. false accept / false reject with verification

- Pros
  - Reflects actual use

- Cons
  - Difficulties /effort in establishing sequence of images per person
  - (without reliance on biometric algorithm)

NPL
National Physical Laboratory

Identification metrics:
Operator viewpoint

Operator's interest

- Specifying and monitoring performance

- Adjustment of parameters to meet operational constraints
    - E.g. fixed resource to deal with number of false alarms per unit of time, and corresponding likelihood of detection
    - Performance achieved depends on **throughput** of subjects, and **probabilities** that subject is on watch-list.

- Metrics of primary interest application specific,
    - Influenced by subject viewpoint metrics.

NPL
National Physical Laboratory

Issues Arising

Establishing ground truth & test identity

- Presence & identity of test subjects
  - both on WL or not on WL

- Correct labelling of segmented images
  - Preferably without referring to biometric algorithm under test
  - Preferably not relying on operator markup

Measuring the extent of a subject's opportunity to be identified

- E.g., subject of interest may be partially obscured by other subjects while notionally "in shot"

- Some subjects visible for a long period, others not at all

- A complex notion of image quality

NPL
National Physical Laboratory

Issues Arising

Attribute errors to different stages of the process?

- Presentation: subject behaviour

- Capture: illumination, camera angle, ...

- Image quality: (of sequence of images of subject)

- Comparison / candidate list properties

- Decision process

  – Automated / or made by human operator (operator GUI?)

Time to update 19795 standards with more guidance on identification testing?

NPL
National Physical Laboratory

Questions?

Contact details

Tony Mansfield

National Physical Laboratory, Teddington, TW11 0LW, UK

+44 20 8943 7029

tony.mansfield@npl.co.uk

NPL
National Physical Laboratory